Evaluating Forecasts for Better Decision-Making in Energy Trading and Risk Management:

An Industry Practitioner's View on How to Enhance the Usefulness of Forecasts Including Potential Applications of Machine Learning

Nazim Osmancik
Chief Risk Officer, Energy Marketing & Trading, Centrica Plc, U.K.

Mr. Nazim Osmancik, Chief Risk Officer, Energy Marketing & Trading, Centrica Plc**,** presenting in Berlin at the Electricity Price Modelling and Forecasting Forum organized by TBM Evolution.

Introduction

Forecasts play a vital role in decision-making in the energy sector as a key input to short-term trading and risk management as well as long-term investment decisions and strategic planning.

The energy transition is bringing new sources of uncertainty such as supply intermittency, demand response, and more volatile spot prices into energy systems that already had a tendency to fall into disequilibrium frequently. This introduces new forecasting challenges. Conversely, as energy systems around the world are transformed and become more dynamic, the commercial importance of having access to accurate forecasts is growing.

This short paper examines some of the key forecasting challenges against this backdrop and introduces ideas on evaluating and enhancing forecasts for better decision-making in the energy trading and risk management context. Case studies are also presented where the ideas are applied and commercial insights or tangible improvements in forecast performance observed.

It is important to note that forecast evaluation and enhancement as well as applications of machine learning to forecasting are large and growing areas of academic research. This article is a non-technical presentation from the perspective of an industry practitioner and does not provide a review of the academic literature. The interested reader is strongly advised to invest time in studying the literature comprehensively.

Emerging Forecasting Challenges in the Energy Sector

One of the key forecasting challenges in our business is the large number of highly variable and interdependent drivers that need to be forecasted. This is due to the fact that energy is a key component of nearly every economic activity and fundamental needs such as heating, which are driven by complex natural systems like weather. Typical energy suppliers or traders, especially those exposed to merchant risk therefore need to understand many complex variables including *inter alia* future commodity prices, market volatility, market positioning of other traders, stocks, weather, maintenance schedules, macroeconomic indicators, FX rates, and so on.

Figure 1
A Large Number of Forecasts are Input to Key Decisions in the Energy Industry

Another challenge comes from the increasing share of intermittent renewable energy and demand side response. The falling cost of renewable energy, ascent of on-site distributed energy, emergence of Internet of Things (IoT), and digitalization of data disrupted the functioning of energy markets designed to serve a predictable future demand load with large centralized generation. In particular, the intermittent nature of renewable generation is creating challenges in system operations as well as driving spot price volatility.

In this environment, the ability to recognize and predict patterns and respond to them in a timely fashion are both harder and more important than ever for maximizing value and managing risks. A typical business would therefore need to have in-house capabilities to produce forecasts or obtain them from external sources to serve their needs.

Apart from being difficult and potentially expensive, this presents yet another challenge, which relates to forecast quality and its variability across forecasters and through time. In the world of energy forecasting, it is not uncommon to find a plus or minus 60% spread around the average forecast for a particular variable, especially if the forecast horizon is longer than a few months. Conversely, for some variables there are very few forecasts. Sometimes the forecast is incomplete or too old.

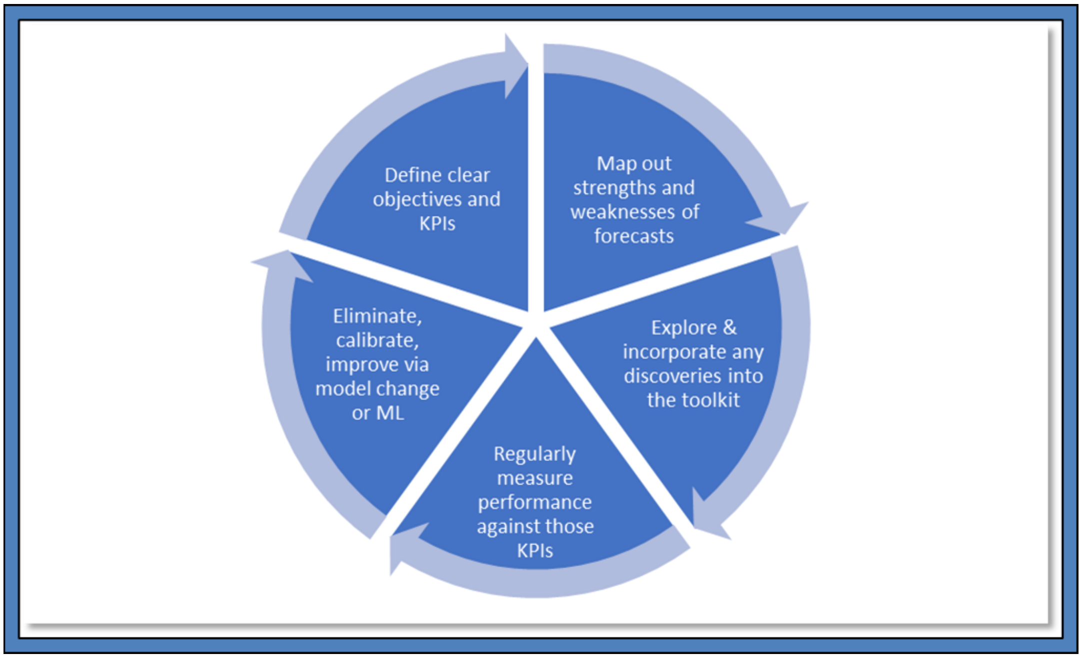
It is of course possible to perform basic modifications such as averaging, taking subsets of our favorite forecasters, or extrapolation. More sophisticated methods that build on Mincer and Zarnowitz (1969) can be used to develop statistical evaluation and backtesting which may be needed to demonstrate adequacy of internal competence. Unfortunately, in practice such methods can prove difficult to connect with commercial or strategic objectives.

It is not all bad news however. While the complexity and scale of the challenges increase, advances in automation, digitalization, predictive techniques including Machine Learning (ML) / Artificial Intelligence (AI) and the ease in implementing them certainly offer new opportunities.

Systematizing Forecast Evaluation and Enhancement

To evaluate the practical usefulness of forecasts and possibly enhance them, developing a systematic approach with a commercial perspective is advisable. This section will cover what such a system might look like and break it down to five processes[1] shown in Figure 2 on the next page.

Figure 2
A Five Step Process of Systematizing Forecast Evaluation and Enhancement



Note: KPI stands for Key Performance Indicator.

Step 1: Defining Objectives and Establishing KPIs

The first step is defining what the business really needs to forecast and what kind of forecast qualities are required. The definition should be as specific as possible. For example, a good definition could look like what is shown in Table 1 below.

Table 1
An Example of How Forecast Requirements Can be Defined

Forecast	Within day movement of the month-ahead futures contract of commodity *X* on day *T*.
Active duration	One day, i.e. forecast to be issued 24 hours before and expire at the end of day *T*.
Primary use	Proprietary trading positions.
Dependencies	Forecasts for commodities *Y*, *Z*; input variables *A, B, C* . . . need to be consistent with this forecast as they are connected.
Desired accuracy	Mean absolute percentage error < *Epsilon*; Hit ratio > *H*
Justification for desired accuracy	Considering transaction costs, VaR limits, counterparties, liquidity, the forecasts within the desired accuracy constraints will generate positive PnL.

Notes: VaR stands for Value-at-Risk, and PnL stands for profit and loss.

Step 2: Understanding the Strengths and Weaknesses of Existing Forecasts and Capabilities

The second step is figuring out whether the forecasts available to the business are adequate using general statistical methods and also with respect to the requirements set out in Step 1. This process should then lead to the identification of any performance gaps.

As part of this process, it is advisable to analyze the behavior and performance of forecasts expansively, e.g., exploring performance in a rising versus falling market, weekdays versus weekends, winters versus summers which may all lead to discoveries that end up being commercially useful. Machine Learning tools such as classification could be very effective for such exploratory tasks.

Step 3: Incorporating Discoveries into the Toolkit

The third step entails identifying "hidden gems" from Step 2 – commercially useful insights carried by forecasts that were unknown and underutilized – and determining what action to take, e.g., allocate risk capital to trade on the insights.

Step 4: Monitoring

The fourth step is building an automation system to monitor the established forecast KPIs with the capability to generate reports, fire signals when performance deteriorates, and integrate into other relevant management information systems.

Step 5: Calibration and Enhancement

In the last step, the rest of the system and data generated can be used to pick the "best" forecasts, or combine and calibrate them to maximize the desired KPIs. As in Step 2, Machine Learning and AI tools can be useful here, though simple econometric methods also perform well.

The impact of this 5-step approach on performance can be significant. The next section will cover some examples where this approach was applied and tangible benefits were observed.

Applications

This section includes a number of examples where publicly available forecasts or forecasts generated by simple econometric models were assessed. Forecasts are anonymized as the purpose of the exercise is to simulate how a generic forecast may be evaluated and enhanced rather than assessing the predictability of a certain market or exploring the capabilities of a forecaster.

Example 1: Understanding the Performance of a Forecast

In this example, it is assumed that the user requires:

1) forecast of a variable that changes through time and expires at a certain date, e.g., a temperature forecast for a future date or a commodity future that expires within a certain number of days;

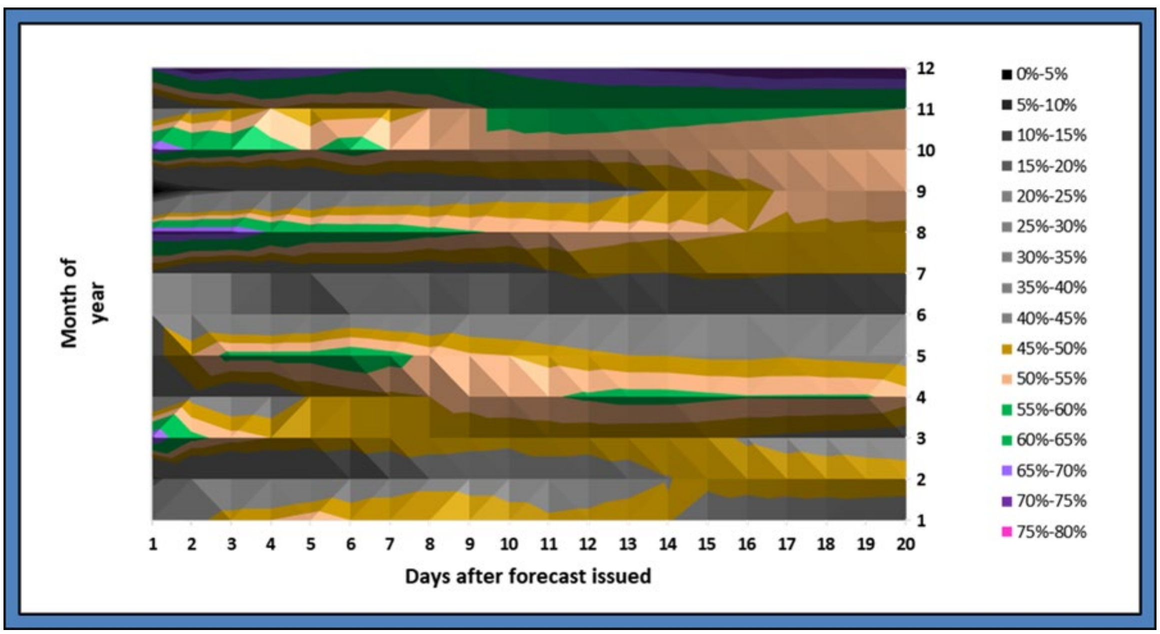
2) prediction of whether the value of the variable will be higher or lower from the time the forecast is published until expiry;

3) prediction accuracy to be higher than 50%, i.e., better than tossing a coin[2]; and

4) an understanding of how forecast performance varies seasonally.

Figure 3 below is an example of a visual that provides pertinent insights. It depicts a forecast's directional accuracy (in predicting whether the variable of interest will rise or fall) for up to 20 days following issuance summarized by month of the year. The sample includes daily data from 2/2/2013 to 5/10/2019 (1724 days).[3]

Figure 3
Forecast Hit Ratio - Number of Days After Issuance by Month of the Year

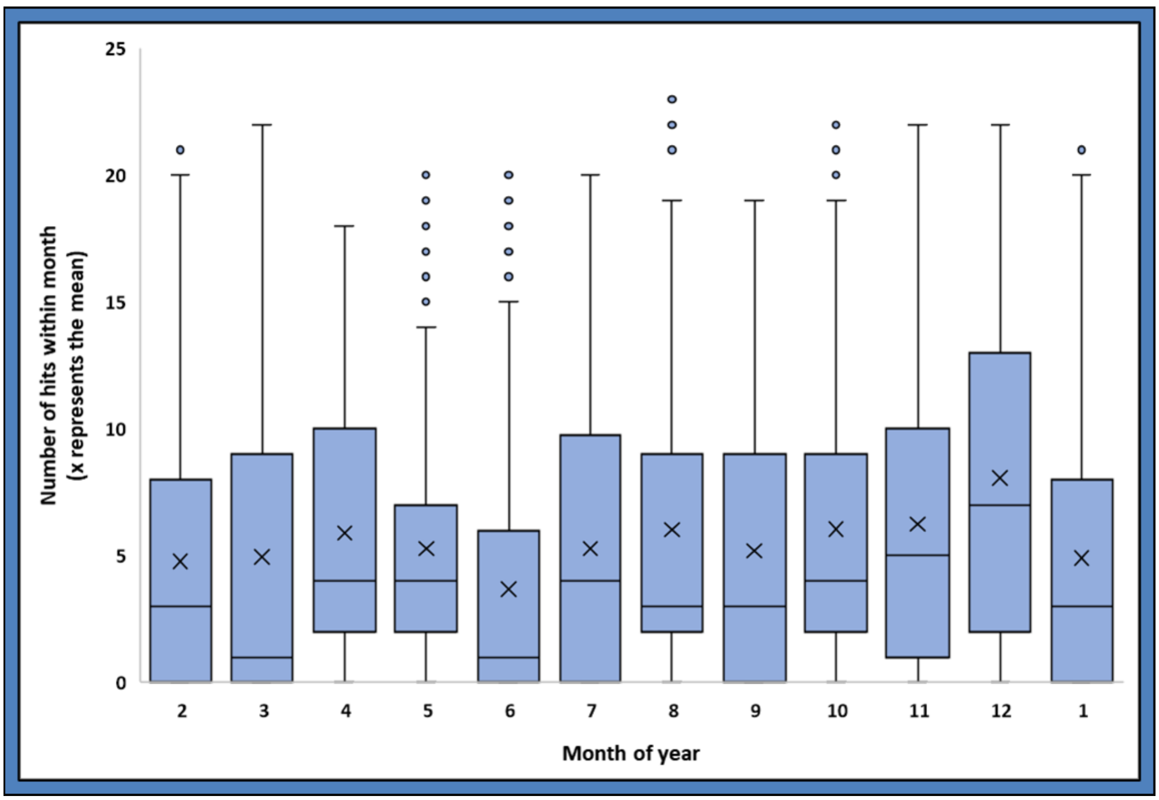


We immediately observe that the forecast tends to perform well in months 4, 5, 8, 10 and 12 but poorly in months 1, 2, 6, 7, and 9. In months 5, 8 and 10 the performance tends to tail off about 9 days after issuance, where in month 3, it does so within 3 days after issuance. Conversely, for months 4, 11, and 12 the forecast attains peak performance 10 days after issuance.

It is difficult to determine whether the pattern is random without deeper knowledge of the underlying process. For example, it could be that in months 6 and 7 quality of input data deteriorates or most forecasters go on vacation which leads to a deterioration of output quality. If the forecasts come from an internal model, this type of analysis can help identify weaknesses and process failures. If the forecasts are sourced from third parties then statistical analysis is required to identify significant patterns.

Using the same data set, Figure 4 then explores the number of days in which a correct prediction was achieved in the format of a box plot to indicate the spread and skewness of the performance by month. One of the key observations here is that within the months where the performance is higher (4, 5, 8, 10, 12), the forecast user will have 5 to 10 opportunities (days) to act on the insight.

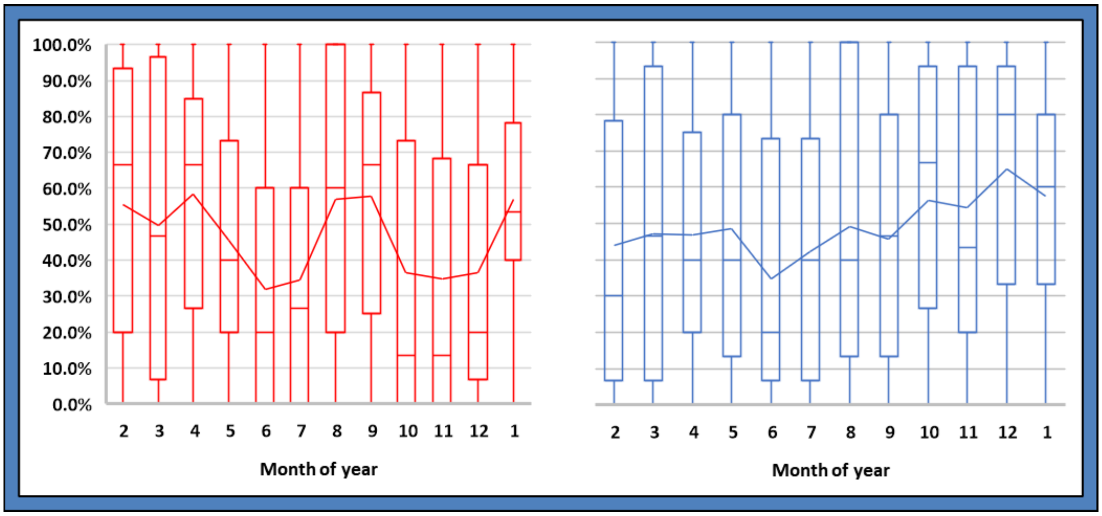
Figure 4
Total Number of Days Within a Month Where the Direction of the Prediction was Correct



Example 2: Enhancing a Forecast

Building on the previous example, we assume the forecast user has similar requirements with access to a forecast that does not seem to perform adequately where the hit ratio is below 50% in summer and winter months as shown on the left-hand side chart in Figure 5 below.

Figure 5
Hit Ratio of Forecast Before (left) and After (right) Adjustment



With no prior knowledge of the underlying process that generated the forecast, historical forecast errors were examined which indicated the forecast was suboptimal as errors exhibited serial correlation as well as episodic periods of bias. Given the lack of information on the underlying process and an obvious theoretical explanation of the biases observed, enhancement was attempted via Machine Learning.

A simple feed-forward neural net was developed to calibrate the forecast using a small number of input variables including the previous day's forecast error, time related variables such as the day of the week, month of the year, and variables that characterized market conditions such as the rolling average of daily volatility. The estimation was set up like a walk-forward backtest where the neural net used historical data to make out-of-sample predictions and was re-estimated regularly as new information became available. No particular care was taken to optimize the size of the neural net or testing the validity of the input variables.

The resulting adjusted forecast performed better under certain conditions (in months 1, 10, 11, 12), attaining higher hit rates, as the chart on the right-hand side of Figure 5 shows. If our hypothetical user was only interested in high performance in winter months, this calibration might have worked well.

More generally, performance of the adjusted forecast was worse than the original forecast in a number of time periods. This often happens in calibrations as improvements come with trade-offs. In this illustrative example, it is likely that the calibration model was not well-specified and could be improved.

Conclusion

The paper examines the key forecasting challenges in the energy sector and introduces a practitioner's approach to understanding, evaluating, and improving forecasts. Simplified use cases are presented, which demonstrate an approach that can generate commercial insights and improvements in forecast performance.

Endnotes

1 In real applications, it is also strongly advisable to define an overarching purpose for implementing this and articulating how it serves a strategic business goal, though this is beyond the scope of this paper.

2 For the sake of simplicity statistical significance requirements are ignored as this is an illustrative example.

3 In order to obtain a smooth picture, a continuous rolling average of the hit ratio has been used.

Disclaimers

Reference

Mincer, J. and V. Zarnowitz, 1969, "The Evaluation of Economic Forecasts," in J. Mincer (Ed) Economic Forecasts and Expectations of Forecasting Behavior and Performance, New York: National Bureau of Economic Research, Inc., pp. 3-46.

Author Biography

NAZIM OSMANCIK
Chief Risk Officer, Energy Marketing & Trading, Centrica Plc, U.K.

Mr. Nazim Osmancik is a senior executive with extensive experience in macro research, strategy, market analysis, trading and risk management gained in the energy sector and professional services. Mr. Osmancik currently leads risk, treasury, foreign exchange and cash management operations in the energy marketing and trading business of Centrica Plc, where he previously led the global market analysis and price forecasting functions. Prior to Centrica, he held various posts in consulting firms including IPA, PwC, and ICFi. Mr. Osmancik studied Economics and Mathematics at Macalester College and has a Master's degree in Finance from the London School of Economics.

His research interests include market pricing in fully decarbonized energy systems, non-linear interactions between energy commodity markets, forecast evaluation and enhancement, and developing systematic trading strategies.